

What is claimed:

1. A computing system comprising a plurality of computing subsystems, each subsystem comprising a processing unit for executing instructions and a local cache memory element for storing a local copy of one or more data elements for high-speed access by said processing unit; and a network switching element comprising a plurality of ports and a storage element, each of said plurality of subsystems being in communication with a different port of said network switching element, said switching element being adapted to monitor transactions transmitted via said ports and generated by said plurality of subsystems, interpret said transactions to determine the status of each of said cache memory elements, and store said status information in said storage element, and route future transactions to a subset of said subsystems based on said stored status information.
2. The system of claim 1, wherein said status information comprises the states of invalid, modified, shared and exclusive for each cache line in each of said local cache memory elements.
3. The system of claim 2, wherein said status information further comprises the state of owner.
4. The system of claim 1, further comprising a shared memory accessible to each of said plurality of subsystems in communication with said switching element.
5. The system of claim 1, wherein said transactions comprise memory read, memory write and cache invalidate operations.
6. The system of claim 1, wherein said subset of subsystems comprises subsystems in which the cache memory contains the data element described in said future transaction.

7. A method of reducing traffic between a plurality of processor subsystems in a distributed processing system through a network switching element, where each of said plurality comprises at least one processing unit and a cache memory element, said method comprising the steps of:
- monitoring all transactions transmitted by said plurality of subsystems to said switching element;
 - interpreting said transactions to deduce the status of said cache memory element in each of said plurality of subsystems;
 - storing said status information associated with each of said cache memory elements in said network switching element; and
 - routing future transactions to a subset of said plurality of subsystems based on said status information.
8. A method of reducing the latency of time critical transmissions through a network switching device, where said latency is defined as the time between receipt of said time critical transmission via a first port and the resending of said time critical transmission via a second port, comprising the steps of:
- receiving a first transmission via said first port;
 - identifying said first transmission as a time critical transmission;
 - sending said first transmission via said second port if said second port is idle;
 - if said second port is not idle, interrupting a second transmission currently in progress via said second port,
 - transmitting a first delimiter to notify recipient of said second transmission that said second transmission is being interrupted,

transmitting said first transmission via said second port,

transmitting a second delimiter to notify recipient of said first and second transmissions that said first transmission has been sent and said second transmission is being resumed, and
transmitting the remainder of said second transmission.